PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04S 5/00 — A1

(11) International Publication Number: WO 95/31881

(43) International Publication Date: 23 November 1995 (23.11.95)

(21) International Application Number: PCT/US95/04839

(22) International Filing Date: 3 May 1995 (03.05.95)

(30) Priority Data:

08/241,867	11 May 1994 (11.05.94)	US
08/303,705	9 September 1994 (09.09.94)	US

(71) Applicant: CRYSTAL RIVER ENGINEERING, INC. [US/US]; Suite 200, 490 California Avenue, Palo Alto, CA 94306 (US).

(72) Inventor: ABEL, Jonathan, S.; 184 Tennyson Avenue, Palo Alto, CA 94301 (US).

(74) Agents: GALLAGHER, Thomas, A. et al.; Law Offices of Thomas A. Gallagher, 3rd floor, 100 Green Street, San Francisco, CA 94111-1302 (US).

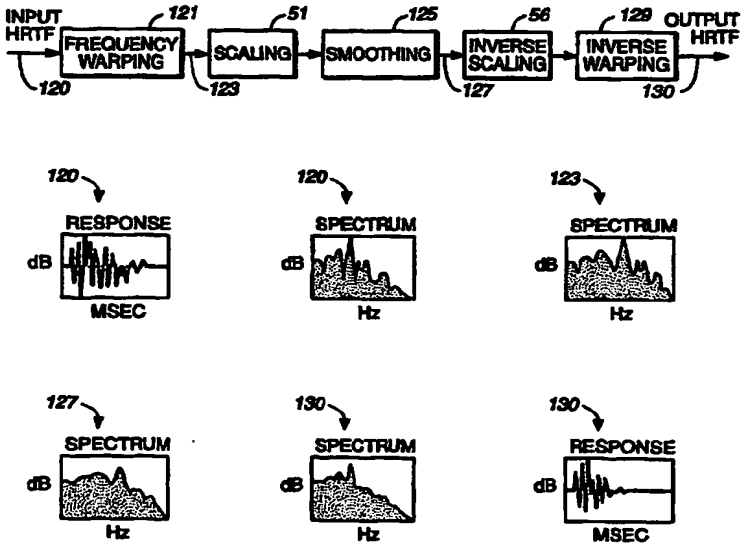
(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THREE-DIMENSIONAL VIRTUAL AUDIO DISPLAY EMPLOYING REDUCED COMPLEXITY IMAGING FILTERS



(57) Abstract

Compressed head-related transfer function (HRTF) (130) parameters are prederived or derived in real time for use in filtering an audio signal for a virtual audio display. From a frequency domain viewpoint, frequency components of known transfer functions are smoothed (125) over bandwidths which are a function of the width of the ear's critical bands. In the first implementation, an HRTF is smoothed (125) by convolving the HRTF (120) with a frequency dependent weighting function in the frequency domain. In the second way, the HRTF's frequency axis is warped or mapped into a non-linear frequency domain.

BNSDOCID: <WO___9531881A1_I_>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

BNSDOCID: <WO___9531881A1_I_>

DESCRIPTION

THREE-DIMENSIONAL VIRTUAL AUDIO DISPLAY EMPLOYING REDUCED COMPLEXITY IMAGING FILTERS

Technical Field

This invention relates generally to three-dimensional or "virtual" audio. More particularly, this invention relates to a method and apparatus for reducing the complexity of imaging filters employed in virtual audio displays. In accordance with the teachings of the invention, such reduction in complexity may be achieved without substantially affecting the psychoacoustic localization characteristics of the resulting three-dimensional audio presentation.

Background Art

Sounds arriving at a listener's ears exhibit propagation effects which depend on the relative positions of the sound source and listener. Listening environment effects may also be present. These effects, including differences in signal intensity and time of arrival, impart to the listener a sense of the sound source location. If included, environmental effects, such as early and late sound reflections, may also impart to the listener a sense of an acoustical environment. By processing a sound so as to simulate the appropriate propagation effects, a listener will perceive the sound to originate from a specified point in three-dimensional space — that is a "virtual" position. See, for example, "Headphone simulation of free-field listening" by Wightman and Kistler, *J. Acoust. Soc. Am.*, Vol. 85, No. 2, 1989.

Current three-dimensional or virtual audio displays are implemented by time-domain filtering an audio input signal with selected head-related transfer functions (HRTFs). Each HRTF is designed to reproduce the propagation effects and acoustic cues responsible for psychoacoustic localization at a particular position or region in three-dimensional space or a direction in three-dimensional space. See, for example, "Localization in Virtual Acoustic Displays" by Elizabeth M. Wenzel, *Presence*, Vol. 1, No. 1, Summer 1992. For simplicity, the present document will refer only to a single

BNSDOCID: <WO___9531881A1_I_>

HRTF operating on a single audio channel. In practice, pairs of HRTFs are employed in order to provide the proper signals to the ears of the listener.

At the present time, most HRTFs are indexed by spatial direction only, the range component being taken into account independently. Some HRTFs define spatial position by including both range and direction and are indexed by position. Although particular examples herein may refer to HRTFs defining direction, the present invention applies to HRTFs representing either direction or position.

HRTFs are typically derived by experimental measurements or by modifying experimentally derived HRTFs. In practical virtual audio display arrangements, a table of HRTF parameter sets are stored, each HRTF parameter set being associated with a particular point or region in three-dimensional space. In order to reduce the table storage requirements, HRTF parameters for only a few spatial positions are stored. HRTF parameters for other spatial positions are generated by interpolating among appropriate sets of HRTF positions which are stored in the table.

As noted above, the acoustic environment may also be taken into account. In practice, this may be accomplished by modifying the HRTF or by subjecting the audio signal to additional filtering simulating the desired acoustic environment. For simplicity in presentation, the embodiments disclosed refer to the HRTFs, however, the invention applies more generally to all transfer functions for use in virtual audio displays, including HRTFs, transfer functions representing acoustic environmental effects and transfer functions representing both head-related transforms and acoustic environmental effects.

A typical prior art arrangement is shown in Figure 1. A three-dimensional spatial location or position signal 10 is applied to an HRTF parameter table and interpolation function 11, resulting in a set of interpolated HRTF parameters 12 responsive to the three-dimensional position identified by signal 10. An input audio signal 14 is applied to an imaging filter 15 whose transfer function is determined by the applied interpolated HRTF parameters. The filter 15 provides a "spatialized" audio output suitable for application to one channel of a headphone 17.

Although the various Figures show headphones for reproduction, appropriate HRTFs may create psychoacoustically localized audio with other types of audio transducers, including loudspeakers. The invention is not limited to use with any particular type of audio transducer.

BNSDOCID: <WO___9531881A1_I_>

When the imaging filter is implemented as a finite-impulse-response (FIR) filter, the HRTF parameters define the FIR filter taps which comprise the impulse response associated with the HRTF. As discussed below, the invention is not limited to use with FIR filters.

The main drawback to the prior art approach shown in Figure 1 is the computational cost of relatively long or complex HRTFs. The prior art employs several techniques to reduce the length or complexity of HRTFs. An HRTF, as shown in Figure 2a, comprises a time delay D component and an impulse response $g(t)$ component. Thus, imaging filters may be implemented as a time delay function z^{-D} and an impulse response function $g(t)$, as shown in Figure 2b. By first removing the time delay, thereby time aligning the HRTFs, the computational complexity of the impulse response function of the imaging filter is reduced.

Figure 3a shows a prior art arrangement in which pairs of unprocessed or "raw" HRTF parameters 100 are applied to a time-alignment processor 101, providing at its outputs time-aligned HRTFs 102 and time-delay values 103 for later use (not shown). Processor 101 cross-correlates pairs of raw HRTFs to determine their time difference of arrival; these time differences are the delay values 103. Because the time delay value values 103 and the filter terms are retained for later use, there is no psychoacoustic localization loss — the perceptual impact is preserved. Each time-aligned HRTF 102 is then processed by a minimum-phase converter 104 to remove residual time delay and to further shorten the time-aligned HRTFs.

Figure 3b shows two left-right pairs (R1/L1 and R2/L2) of exemplary raw HRTFs resulting from raw HRTF parameters 100. Figure 3c shows corresponding time-aligned HRTFs 102. Figure 3d shows the corresponding output minimum-phase HRTFs 105. The impulse response lengths of the time-aligned HRTFs 102 are shortened with respect to the raw HRTFs 100 and the minimum-phase HRTFs 105 are shortened with respect to the time-aligned HRTFs 102. Thus, by extracting the delay so as to time align the HRTFs and by applying minimum phase conversion, the filter complexity (its length, in the case of an FIR filter) is reduced.

Despite the use of the techniques of Figures 2b and 3a, at an audio sampling rate of 48 kHz, minimum phase responses as long as 256 points for an FIR filter are commonly used, requiring processors executing on the order of 25 mips per audio source rendered.

BNSDOCID: <WO___9531881A1_I_>

When computational resources are limited, two additional approaches are used in the prior art, either singly or in combination, to further reduce the length or complexity of HRTFs. One technique is to reduce the sampling rate by down sampling the HRTF as shown in Figure 4a. Since many localization cues, particularly those important to elevation, involve high-frequency components, reducing the sampling rate may unacceptably degrade the performance of the audio display.

Another technique, shown in Figure 4b, is to apply a windowing function to the HRTF by multiplying the HRTF by a windowing function in the time domain or by convolving the HRTF with a corresponding weighting function in the frequency domain. This process is most easily understood by considering the multiplication of the HRTF by a window in the time domain — the window width is selected to be narrower than the HRTF, resulting in a shortened HRTF. Such windowing results in a frequency-domain smoothing with a fixed weighting function. This known windowing technique degrades psychoacoustic localization characteristics, particularly with respect to spatial positions or directions having complex or long impulse responses. Thus, there is a need for a way to reduce the complexity or length of HRTFs while maintaining the perceptual impact and psychoacoustic localization characteristics of the original HRTFs.

Disclosure of Invention

In accordance with the present invention, a three-dimensional virtual audio display generates a set of transfer function parameters in response to a spatial location signal and filters an audio signal in response to the set of head-related transfer function parameters. The set of head-related transfer function parameters are smoothed versions of parameters for known head-related transfer functions.

The smoothing according to the present invention is best explained by considering its action in the frequency domain: the frequency components of known transfer functions are smoothed over bandwidths which are a non-constant function of frequency. The parameters of the resulting transfer functions, referred to herein as "compressed" transfer functions, are used to filter the audio signal for the virtual audio display. The compressed head-related transfer function parameters may be prederived or may be derived in real time. Preferably, the smoothing bandwidth is a function of the width of the ear's critical bands (i.e., a function of "critical bandwidth"). The function may be such that the smoothing bandwidth is proportional to critical bandwidth. As is well

BNSDOCID: <WO___9531881A1_I_>

known, the ear's critical bands increase in width with increasing frequency, thus the smoothing bandwidth also increases with frequency.

The wider the smoothing bandwidth relative to the critical bandwidth, the less complex the resulting HRTF. In the case of an HRTF implemented as an FIR filter, the length of the filter (the number of filter taps) is inversely related to the smoothing bandwidth expressed as a multiple of critical bandwidth.

By applying the teachings of the present invention which take critical bandwidth into account, for the same reduction in complexity or length, the resulting less complex or shortened HRTFs have less degradation of perceptual impact and psychoacoustic localization than HRTFs made less complex or shortened by prior art windowing techniques such as described above.

An example HRTF ("raw HRTF") and shortened versions produced by a prior art windowing method ("prior art HRTF") and by the method according to the present invention ("compressed HRTF") are shown in Figures 5a (time domain) and 5b (frequency domain). The raw HRTF is an example of a known HRTF that has not been processed to reduce its complexity or length. In Figure 5a, the HRTF time-domain impulse response amplitudes are plotted along a time axis of 0 to 3 milliseconds. In Figure 5b the frequency-domain transfer function power of each HRTF is plotted along a log frequency axis extending from 1 kHz to 20 kHz. In the time domain, Figure 5a, the prior art HRTF exhibits some shortening, but the compressed HRTF exhibits even more shortening. In the frequency domain, Figure 5b, the effect of uniform smoothing bandwidth on the prior art HRTF is apparent, whereas the compressed HRTF shows the effect of an increasing smoothing bandwidth as frequency increases. Because of the log frequency scale of Figure 5b, the compressed HRTF displays a constant smoothing with respect to the raw HRTF. Despite their differences in time-domain length and frequency-domain frequency response, the raw HRTF, the prior art HRTF, and the compressed HRTF provide comparable psychoacoustic performance.

When the amount of prior art windowing and compression according to the present invention are chosen so as to provide substantially similar psychoacoustic performance with respect to raw HRTFs, preliminary double-blind listening tests indicate a preference for compressed HRTFs over prior art windowed HRTFs. Somewhat surprisingly, compressed HRTFs were also preferred over raw HRTFs. This is believed to be

BNSDOCID: <WO___9531881A1_I_>

because the HRTF fine structure eliminated by the smoothing process is uncorrelated from HRTF position to HRTF position and may be perceived as a form of noise.

The present invention may be implemented in at least two ways. In a first way, an HRTF is smoothed by convolving the HRTF with a frequency dependent weighting function in the frequency domain. This weighting function differs from the frequency domain dual of the prior art time-domain windowing function in that the weighting function varies as a function of frequency instead of being invariant. Alternatively, a time-domain dual of the frequency dependent weighting function may be applied to the HRTF impulse response in the time domain. In a second way, the HRTF's frequency axis is warped or mapped into a non-linear frequency domain and the frequency-warped HRTF is either multiplied by a conventional window function in the time domain (after transformation to the time domain) or convolved with the non-varying frequency response of the conventional window function in the frequency domain. Inverse frequency warping is subsequently applied to the windowed signal.

The present invention may be implemented using any type of imaging filter, including, but not limited to, analog filters, hybrid analog/digital filters, and digital filters. Such filters may be implemented in hardware, software or hybrid hardware/software arrangements, including, for example, digital signal processing. When implemented digitally or partially digitally, FIR, IIR (infinite-impulse-response) and hybrid FIR/IIR filters may be employed. The present invention may also be implemented by a principal component filter architecture. Other aspects of the virtual audio display may be implemented using any combination of analog, digital, hybrid analog/digital, hardware, software, and hybrid hardware/software techniques, including, for example, digital signal processing.

In the case of an FIR filter implementation, the HRTF parameters are the filter taps defining the FIR filter. In the case of an IIR filter, the HRTF parameters are the poles and zeroes or other characteristics defining the IIR filter. In the case of a principal component filter, the HRTF parameters are the position-dependent weights.

In another aspect of the invention, each HRTF in a group of HRTFs is split into a fixed head-related transfer function common to all head-related transfer functions in the group and a variable head-related transfer function associated with respective head-related transfer functions, the combination of the fixed and each variable head-related transfer function being substantially equivalent to the respective original known head-

BNSDOCID: <WO___9531881A1_I_>

related transfer function. The smoothing techniques according to the present invention may be applied to either the fixed HRTF, the variable HRTF, to both, or to neither of them.

Brief Description of Drawings

Figure 1 is a functional block diagram of a prior art virtual audio display arrangement.

Figure 2a is an example of the impulse response of a head-related transfer function (HRTF).

Figure 2b is a functional block diagram illustrating the manner in which an imaging filter may represent the time-delay and impulse response portions of an HRTF.

Figure 3a is a functional block diagram of one prior art technique for reducing the complexity or length of an HRTF.

Figure 3b is a set of example left and right "raw" HRTF pairs.

Figure 3c is the set of HRTF pairs as in Figure 3b which are now time aligned to reduce their length.

Figure 3d is the set of HRTF pairs as in Figure 3c which are now minimum phase converted to further reduce their length.

Figure 4a is a functional block diagram showing a prior art technique for shortening an HRTF impulse response by reducing the sampling rate.

Figure 4b is a functional block diagram showing a prior art technique for shortening an HRTF impulse response by multiplying it by a window in the time domain.

Figure 5a is a set of three waveforms in the time domain, illustrating an example of a "raw" HRTF, the HRTF shortened by prior art techniques and the HRTF compressed according to the teachings of the present invention.

Figure 5b is a frequency domain representation of the set of HRTF waveforms of Figure 5a.

Figure 6a is a functional block diagram showing an embodiment for deriving compressed HRTFs according to the present invention.

Figure 6b shows the frequency response of an exemplary input HRTF.

Figure 6c shows the impulse response of the exemplary input HRTF impulse response.

Figure 6d shows the frequency response of the compressed output HRTF.

BNSDOCID: <WO___9531881A1_I_>

Figure 6e shows the impulse response of the compressed output HRTF.

Figure 7a shows an alternative embodiment for deriving compressed HRTFs according to the present invention.

Figure 7b shows the impulse response of an exemplary input HRTF impulse response.

Figure 7c shows the frequency response of the exemplary input HRTF.

Figure 7d shows the frequency response of the input HRTF after frequency warping.

Figure 7e shows the frequency response of the compressed output HRTF.

Figure 7f shows the frequency response of the compressed output HRTF after inverse frequency warping.

Figure 7g shows the impulse response of the compressed output HRTF after inverse frequency warping.

Figure 8 shows three of a family of windows useful in understanding the operation of the embodiments of Figures 6a and 7a.

Figure 9 is a functional block diagram in which the imaging filter is embodied as a principal component filter.

Figure 10 is a functional block diagram showing another aspect of the present invention.

Modes for Carrying Out the Invention

Figure 6a shows an embodiment for deriving compressed HRTFs according to the present invention. According to this embodiment, an input HRTF is smoothed by convolving the frequency response of the input HRTF with a frequency dependent weighting function in the frequency domain. Alternatively, a time-domain dual of the frequency dependent weighting function may be applied to the HRTF impulse response in the time domain.

Figure 7a shows an alternative embodiment for deriving compressed HRTFs according to the present invention. According to this embodiment, the frequency axis of the input HRTF is warped or mapped into a non-linear frequency domain and the frequency-warped HRTF is convolved with the frequency response of a non-varying weighting function in the frequency domain (a weighting function which is the dual of a conventional time-domain windowing function). Inverse frequency warping is then

BNSDOCID: <WO___9531881A1_I_>

applied to the smoothed signal. Alternatively, the frequency-warped HRTF may be transformed into the time domain and multiplied by a conventional window function.

Referring to Figure 6a, an optional nonlinear scaling function 51 is applied to an input HRTF 50. A smoothing function 54 is then applied to the HRTF 52. If nonlinear scaling is applied to the input HRTF, an inverse scaling function 56 is then applied to the smoothed HRTF 54. A compressed HRTF 57 is provided at the output. As explained further below, the nonlinear scaling 51 and inverse scaling 56 can control whether the smoothing mean function is with respect to signal amplitude or power and whether it is an arithmetic averaging, a geometric averaging or another mean function.

The smoothing processor 54 convolves the HRTF with a frequency-dependent weighting function. The smoothing processor may be implemented as a running weighted arithmetic mean,

$$S(f) = \frac{1}{2b_f + 1} \sum_{n=-b_f}^{b_f} W_f(n) \cdot H(f-n),$$

Equation 1.

where at least the smoothing bandwidth b_f and, optionally, the window shape W_f are a function of frequency. The width of the weighting function increases with frequency; preferably, the weighting function length is a multiple of critical bandwidth: the shorter the required HRTF impulse response length, the greater the multiple.

HRTFs typically lack low-frequency content (below about 300 Hz) and high-frequency content (above about 16 kHz). In order to provide the shortest possible (and, hence, least complex) HRTFs, it is desirable to extend HRTF frequency response to or even beyond the normal lower and upper extremes of human hearing. However, if this is done, the width of the weighting function in the extended low-frequency and high-frequency audio-band regions should be wider relative to the ear's critical bands than the multiple of critical bandwidth used through the main, unextended portion of the audio band in which HRTFs typically have content.

Below about 500 Hz, HRTFs are approximately flat spectrally because audio wavelengths are large compared to head size. Thus, a smoothing bandwidth wider than the above-mentioned multiple of critical bandwidth preferably is used. At high frequencies, above about 16 kHz, a smoothing bandwidth wider than the above-mentioned multiple of critical bandwidth preferably is also used because human hearing

is poor at such high frequencies and most localization cues are concentrated below such high frequencies. Thus, the weighting bandwidth at the low-frequency and high-frequency extremes of the audio band preferably may be widened beyond the bandwidths predicted by the equations set forth herein. For example, in one practical embodiment of the invention, a constant smoothing bandwidth of about 250 Hz is used for frequencies below 1 kHz, and a third-octave bandwidth is used above 1 kHz. One-third octave bandwidth approximates critical bandwidth; at 1 kHz the one-third octave bandwidth is about 250 Hz. Thus, below 1 kHz the smoothing bandwidth is wider than the critical bandwidth. In some cases, power noted at low frequencies (say, in the range 300 to 500 Hz) is extrapolated to DC to fill in data not accurately determined using conventional HRTF measurement techniques.

Although a weighting function having the same multiple of critical bandwidth may be used in processing all of the HRTFs in a group, weighting functions having different critical bandwidth multiples may be applied to respective HRTFs so that not all HRTFs are compressed to the same extent — this may be necessary in order to assure that the resulting compressed HRTFs are generally of the same complexity or length (certain ones of the raw HRTFs will be of greater complexity or length depending on the spatial location which they represent and may therefore require greater or lesser compression). Alternatively, HRTFs representing certain directions or spatial positions may be compressed less than others in order to maintain the perception of better overall spatial localization while still obtaining some overall lessening in computational complexity. The amount of HRTF compression may be varied as a function of the relative psychoacoustic importance of the HRTF. For example, early reflections, which are rendered using separate HRTFs because they arrive from different directions, are not as important to spatialize as accurately as is the direct sound path. Thus, early reflections could be rendered using "over shortened" HRTFs without perceptual impact.

Another way to view the smoothing 54 of Figure 6a is that for each frequency f,

$$S_\theta(f) = \sum_{n=0}^{N} W_{f,\theta}(n) \cdot H_\theta(n), \qquad \text{Equation 2.}$$

$$\text{where } \sum_{n=0}^{n=N} W_{f,\theta}(n)=1, \qquad \text{Equation 3.}$$

$$W_{f,\theta}(n) \geq 0, \text{ for all } n, \qquad \text{Equation 4.}$$

BNSDOCID: <WO___9531881A1_I_>

$H_\theta(n)$ is the input HRTF 52 at position Θ, $S_\theta(f)$ is the compressed HRTF 54, n is frequency, and N is one half the Nyquist frequency. Thus, there are a family of weighting functions $W_{f,\theta}(n)$, each defined on an interval 0 to N, which have a width which is a function of their center frequency f and, optionally, also a function of the HRTF position Θ. The summation of each weighting function is 1 (Equation 3). Figure 8 shows three members of a family of Gaussian-shaped weighting functions with their amplitude response plotted against frequency. Only three of the family of weighting functions are shown for simplicity. The center window is centered at frequency n_0 and has a bandwidth $b_{f=n_0}$. The weighting functions need not have a Gaussian shape. Other shaped weighting functions, including rectangular, for simplicity, may be employed. Also, the weighting functions need not be symmetrical about their center frequency.

Taking into account the nonlinear scaling function 51 and the inverse scaling function 56, Figure 6a may be more generally characterized as

$$S_\theta(f) = G^{-1}\left\{\sum_{n=0}^{N} W_{f,\theta}(n)\cdot G\{H_\theta(n)\}\right\} \qquad \text{Equation-5.}$$

where G is the scaling 51 and G^{-1} is the inverse scaling.

While the smoothing 54 thus far described provides an arithmetic mean function, depending on the statistics of the input HRTF transfer function, a trimmed mean or median might be favored over the arithmetic mean.

Because the human ear appears to be sensitive to the total filter power in a critical band, it is preferred to implement the nonlinear scaling 51 of Figure 6a as a magnitude squared operation and the output inverse scaler 56 as a square root. It may be desirable to apply certain pre-processing or post-processing such as minimum phase conversion. Alternatively, or in addition to the magnitude squared scaling and square root inverse scaling, the arithmetic mean of the smoothing 54 becomes a geometric mean when the nonlinear scaling 51 provides a logarithm function and the inverse scaling 56 an exponentiation function. Such a mean is useful in preserving spectral nulls thought to be important for elevation perception.

Figures 6b and 6c show an exemplary input HRTF frequency spectrum and input impulse response, respectively, in the frequency domain and the time domain. Figures 6d and 6e show the compressed output HRTF 57 in the respective domains. The degree to which the HRTF spectrum is smoothed and its impulse response is shortened will

BNSDOCID: <WO___9531881A1_I_>

depend on the multiple of critical bandwidth chosen for the smoothing 54. The compressed HRTF characteristics will also depend on the window shape and other factors discussed above.

Refer now to Figure 7a. In this embodiment the frequency axis of the input HRTF is altered by a frequency warping function 121 so that a constant-bandwidth smoothing 125 acting on the warped frequency spectrum implements the equivalent of smoothing 54 of Figure 6a. The smoothed HRTF is processed by an inverse warping 129 to provide the output compressed HRTF. In the same manner as in Figure 6a, nonlinear scaling 51 and inverse scaling 56 optionally may be applied to the input and output HRTFs.

The frequency warping function 121 in conjunction with constant bandwidth smoothing serves the purpose of the frequency-varying smoothing bandwidth of the Figure 6a embodiment. For example, a warping function mapping frequency to Bark may be used to implement critical-band smoothing. Smoothing 125 may be implemented as a time-domain window function multiplication or as a frequency-domain weighting function convolution similar to the embodiment of Figure 6a except that the weighting function width is constant with frequency. As with respect to Figure 6a, it may be desirable to apply certain pre-processing or post-processing such as minimum phase conversion.

The order in which the frequency warping function 121 and the scaling function 51 are applied may be reversed. Although these functions are not linear, they do commute because the frequency warping 121 affects the frequency domain while the scaling 51 affects only the value of the frequency bins. Consequently, the inverse scaling function 56 and the inverse warping function 129 may also be reversed.

As a further alternative, the output HRTF may be taken after block 125, in which case inverse scaling and inverse warping may be provided in the apparatus or functions which receive the compressed HRTF parameters.

Figures 7b and 7c show an exemplary input HRTF input response and frequency spectrum, respectively. Figure 7d shows the frequency spectrum of the HRTF mapped into Bark. Figure 7e shows the spectrum of the HRTF after smoothing 125. After undergoing inverse frequency warping, the resulting compressed HRTF has a spectrum as shown in Figure 7f and an impulse response as shown in Figure 7g. It will be noted

BNSDOCID: <WO__9531881A1_I_>

that the resulting HRTF characteristics are the same as those of the embodiment of Figure 6a.

The imaging filter may also be embodied as a principal component filter in the manner of Figure 9. A position signal 30 is applied to a weight table and interpolation function 31 which is functionally similar to block 11 of Figure 1. The parameters provided by block 31, the interpolated weights, the directional matrix and the principal component filters are functionally equivalent to HRTF parameters controlling an imaging filter. The imaging filter 15' of this embodiment filters the input signal 33 in a set of parallel fixed filters 34, principal component filters, PC_0 through PC_N, whose outputs are mixed via a position-dependent weighting to form an approximation to the desired imaging filter. The accuracy of the approximations increase with the number of principal component filters used. More computational resources, in the form of additional principal component filters, are needed to achieve a given degree of approximation to a set of raw HRTFs than to versions compressed in accordance with this embodiment of the present invention.

Another aspect of the invention is shown in the embodiment of Figure 10. A three-dimensional spatial location or position signal 70 is applied to an equalized HRTF parameter table and interpolation function 71, resulting in a set of interpolated equalized HRTF parameters 72 responsive to the three-dimensional position identified by signal 70. An input audio signal 73 is applied to an equalizing filter 74 and an imaging filter 75 whose transfer function is determined by the applied interpolated equalized HRTF parameters. Alternatively, the equalizing filter 74 may be located after the imaging filter 75. The filter 75 provides a spatialized audio output suitable for application to one channel of a headphone 77.

The sets of equalized head-related transfer function parameters in the table 71 are prederived by splitting a group of known head-related transfer functions into a fixed head-related transfer function common to all head-related transfer functions in the group and a variable, position-dependent head-related transfer function associated with each of the known head-related transfer functions, the combination of the fixed and each variable head-related transfer function being substantially equal to the respective original known head-related transfer function. The equalizing filter 74 thus represents the fixed head-related transfer function common to all head-related transfer functions in the table. In this manner the HRTFs and imaging filter are reduced in complexity.

BNSDOCID: <WO___9531881A1_I_>

The equalization filter characteristics are chosen to minimize the complexity of the imaging filters. This minimizes the size of the equalized HRTF table, reduces the computational resources for HRTF interpolation and image filtering and reduces memory resources for tabulated HRTFs. In the case of FIR imaging filters, it is desired to minimize filter length.

Various optimization criteria may be used to find the desired equalization filter. The equalization filter may approximate the average HRTF, as this choice makes the position-dependent portion spectrally flat (and short in time) on average. The equalization filter may represent the diffuse field sound component of the group of known transfer functions. When the equalization filter is formed as a weighted average of HRTFs, the weighting should give more importance to longer or more complex HRTFs.

Different fixed equalization may be provided for left and right channels (either before or after the position variable HRTFs) or a single equalization may be applied to the monaural source signal (either as a single filter before the monaural signal is split into left and right components or as two filters applied to each of the left and right components). As might be expected from human symmetry, the optimal left-ear and right-ear equalization filters are often nearly identical. Thus, the audio source signal may be filtered using a single equalization filter, with its output passed to both position-dependent HRTF filters.

Further benefits may be achieved by smoothing either the equalized HRTF parameters, the parameters of the fixed equalizing filter or both the equalized HRTF parameters and equalizing filter parameters in accordance with the teachings of the present invention.

Also, using different filter structures for the equalization filter and the imaging filter may result in computational savings: for example, one may be implemented as an IIR filter and the other as an FIR filter. Because it is a fixed filter typically with a fairly smooth response, the equalizing filter may best be implemented as a low-order IIR filter. Also, it could readily be implemented as an analog filter.

Any filtering technique appropriate for use in HRTF filters, including principal component methods, may be used to implement the variable, position-dependent portion equalized HRTF parameters. For example, Figure 10 may be modified to employ as

BNSDOCID: <WO___9531881A1_I_>

imaging filter 75 a principal component imaging filter 15' of the type described in connection with the embodiment of Figure 9.

BNSDOCID: <WO___9531881A1_I_>

CLAIMS

1. A three-dimensional virtual audio display method comprising

generating a set of transfer function parameters in response to a spatial location or direction signal, and

filtering an audio signal in response to said set of transfer function parameters, wherein said set of transfer function parameters selected from or interpolated among parameters is derived by

smoothing frequency components of a known transfer function over a bandwidth which is a non-constant function of frequency, and

noting the parameters of the transfer function of the resulting compressed transfer function.

2. An audio display method according to claim 1 wherein the bandwidth is a function of critical bandwidth.

3. An audio display method according to claim 2 wherein the smoothing comprises, for each frequency component in at least part of the audio band of the display, applying a mean function to the frequency components within the bandwidth containing the frequency component.

4. An audio display method according to claim 3 wherein the mean function is a function of the amplitude of the frequency components.

5. An audio display method according to claim 3 wherein the mean function is a function of the power of the frequency components.

6. An audio display method according to claim 4 or claim 5 wherein said mean function determines the median.

7. An audio display method according to claim 4 or claim 5 wherein said mean function determines the weighted arithmetic mean.

BNSDOCID: <WO___9531881A1_I_>

8. An audio display method according to claim 4 or claim 5 wherein said mean function determines the weighted geometric mean.

9. An audio display method according to claim 4 or claim 5 wherein said mean function determines a trimmed mean.

10. An audio display method according to claim 2 wherein said weighting function has a rectangular shape.

11. An audio display method according to claim 1 wherein the bandwidth is proportional to critical bandwidth.

12. An audio display method according to claim 11 wherein said transfer function parameters are extended at low and high frequencies and wherein said bandwidth is wider than a bandwidth proportional to critical bandwidth in said low- and high-frequency regions.

13. An audio display method according to claim 1 wherein the smoothing comprises convolving the transfer function with a frequency dependent weighting function, the width of which is a function of critical bandwidth.

14. An audio display method according to claim 13 wherein the weighting function has a bandwidth which is a multiple (one or greater) of critical bandwidth.

15. An audio display method according to claim 14 wherein said transfer function parameters are extended at low and high frequencies and wherein said bandwidth is wider than a bandwidth proportional to critical bandwidth in said low- and high-frequency regions.

16. An audio display method according to claim 13 wherein said weighting function has a shape having a higher-order continuity than a rectangularly-shaped window.

BNSDOCID: <WO___9531881A1_I_>

17. An audio display method according to claim 1 wherein smoothing frequency components comprises smoothing said frequency components in the frequency domain.

18. An audio display method according to claim 17 wherein said smoothing comprises convolving said known transfer function $H(f)$ with the frequency response of a weighting function $w_f(i)$ in the frequency domain according to the relationship

$$S(f) = \frac{1}{2b_f + 1} \sum_{n=-b_f}^{b_f} W_f(n) \cdot H(f-n),$$

where at least the smoothing bandwidth b_f and, optionally, the weighting function shape W_f are a function of frequency.

19. An audio display method according to claim 1 wherein smoothing frequency components comprises applying a frequency warping function to said known transfer function, transforming the frequency-warped transfer function to the time domain, and time-domain windowing the impulse response of the frequency-warped transfer function.

20. An audio display method according to claim 1 wherein smoothing frequency components comprises applying a frequency warping function to said known transfer function and frequency-domain convolving the frequency-warped transfer function with the frequency response of a constant weighting function.

21. An audio display method according to claim 19 or claim 20 wherein said frequency warping function maps the transfer function to Bark.

22. An audio display method according to claim 19 or claim 20 further comprising applying a non-linear scaling to said known transfer function prior to said multiplication or said convolving and applying an inverse scaling to the windowed or convolved transfer function.

BNSDOCID: <WO___9531881A1_I_>

23. An audio display method according to claim 1 wherein said filtering is principal-component filtering.

24. An audio display method according to claim 1 wherein said transfer function parameters are equalized transfer function parameters and said filtering includes fixed equalization filtering and filtering in response to said equalized transfer function parameters.

25. An audio display method according to claim 1 wherein said set of transfer functions are derived by smoothing frequency components of known transfer functions over different bandwidths as a function of the spatial location or directions associated with the transfer function.

26. An audio display method according to claim 1 wherein said set of transfer functions are derived by smoothing frequency components of known transfer functions over different bandwidths as a function of the complexity of the transfer function.

27. An audio display method according to claim 1 wherein said set of transfer functions are derived by smoothing frequency components of known transfer functions over different bandwidths as a function of the spatial location or direction associated with the transfer function and as a function of the complexity of the transfer function.

28. An audio display method according to claim 26 or 27 wherein the bandwidth increases with increasing transfer function complexity.

29. An audio display method according to claim 1 or claim 28 wherein the bandwidth is selected such that the most complex resulting compressed transfer function does not exceed a predetermined complexity.

30. An audio display method according to claim 1 wherein said set of transfer functions are derived by smoothing frequency components of known transfer functions over different bandwidths as a function of the relative psychoacoustic importance of the transfer function.

BNSDOCID: <WO___9531881A1_I_>

31. An audio display method according to claim 1 wherein said set of transfer functions are derived by smoothing frequency components of known transfer functions over different bandwidths as a function of the spatial location or direction associated with the transfer function and as a function of the relative psychoacoustic importance of the transfer function.

32. A three-dimensional virtual audio display method comprising

generating a set of equalized transfer function parameters in response to a spatial location or direction signal, and

filtering an audio signal with fixed equalization filtering and in response to said set of equalized transfer function parameters,

wherein said fixed equalization filtering are derived by and said set of equalized transfer function parameters are selected from or interpolated among parameters derived by

splitting a group of known transfer functions into a fixed transfer function common to all transfer functions in the group and a variable transfer function associated with each of the known transfer functions, the combination of the fixed and each variable transfer function being substantially equal to the respective original known transfer function,

noting the parameters of said fixed transfer function for characterizing said fixed equalization filtering, and

noting the parameters of each of the transfer functions of the resulting variable transfer function for use as said equalized transfer function parameters.

33. An audio display method according to claim 28 wherein the derivation of said fixed equalization filtering and said set of equalized transfer function parameters further includes

smoothing frequency components of each of the variable transfer functions over a bandwidth which is a non-constant function of frequency.

34. An audio display method according to claim 28 wherein the derivation of said fixed equalization filtering and said set of equalized transfer function parameters further includes

BNSDOCID: <WO___9531881A1_I_>

smoothing frequency components of the fixed transfer function over a bandwidth which is a non-constant function of frequency.

35. An audio display method according to claim 28 wherein said group of known transfer functions is split into a fixed transfer function and a plurality of variable transfer functions by selecting a fixed transfer function resulting in the least complex variable transfer functions.

36. An audio display method according to claim 28 wherein said group of known transfer functions is split into a fixed transfer function and a plurality of variable transfer functions by selecting a fixed transfer function representing the diffuse field sound component of the group of known transfer functions.

37. An audio display method according to claim 28 wherein said group of known transfer functions are transfer functions representing a particular direction or range of directions in space.

38. An audio display method according to claim 28 comprising the additional step of smoothing frequency components of the fixed transfer function over a bandwidth which is a non-constant function of frequency and wherein the step of noting the parameters of said fixed transfer function for characterizing said fixed equalization filtering notes the parameters of the resulting compressed fixed transfer function.

39. An audio display method according to claim 28 wherein sets of equalized transfer function parameters generated in response to a spatial location or direction signal are generated by principal-component filtering.

40. Three-dimensional virtual audio display apparatus comprising

means for generating a set of transfer function parameters in response to a spatial location or direction signal, said parameters selected from or interpolated among parameters obtained by

smoothing frequency components of a known transfer function over a bandwidth which is a non-constant function of frequency, and

BNSDOCID: <WO___9531881A1_I_>

noting the parameters of the transfer function of the resulting compressed transfer function, and

means for filtering an audio signal in response to said set of transfer function parameters.

41. A three-dimensional virtual audio display method comprising

means for generating a set of equalized transfer function parameters in response to a spatial location or direction signal, said parameters selected from or interpolated among parameters obtained by

splitting a group of known transfer functions into a fixed transfer function common to all transfer functions in the group and a variable transfer function associated with each of the known transfer functions, the combination of the fixed and each variable transfer function being substantially equal to the respective original known transfer function,

noting the parameters of said fixed transfer function for characterizing said fixed equalization filtering, and

noting the parameters of each of the transfer functions of the resulting variable transfer function for use as said equalized transfer function parameters, and

means for filtering an audio signal with fixed equalization filtering and in response to said set of equalized transfer function parameters.

BNSDOCID: <WO__9531881A1_I_>

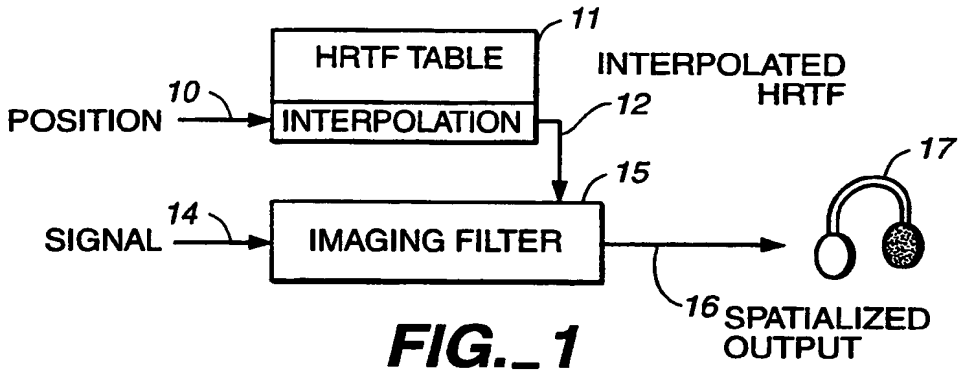


FIG._1

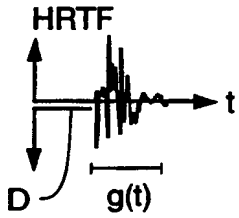


FIG._2a

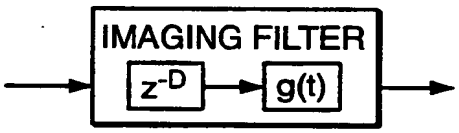


FIG._2b

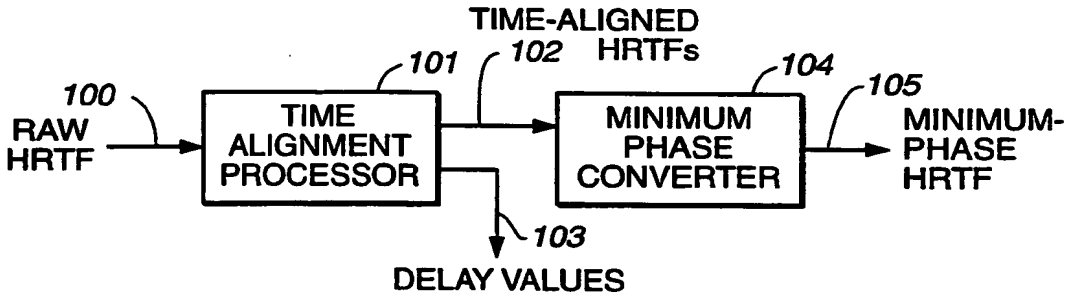


FIG._3a

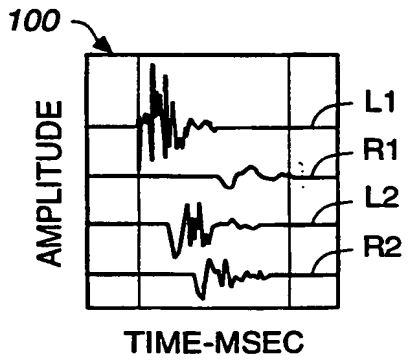


FIG._3b

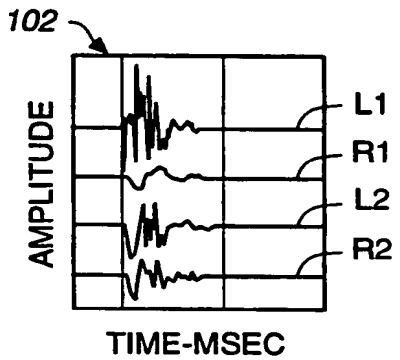


FIG._3c

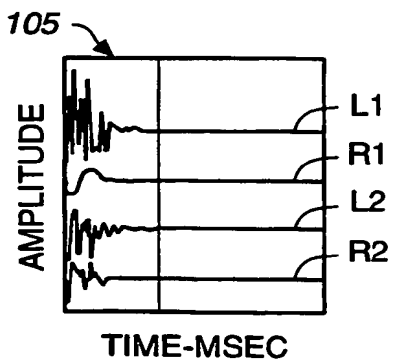


FIG._3d

BNSDOCID: <WO___9531881A1_I_>

FIG._4a

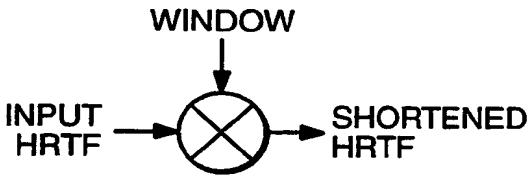


FIG._4b

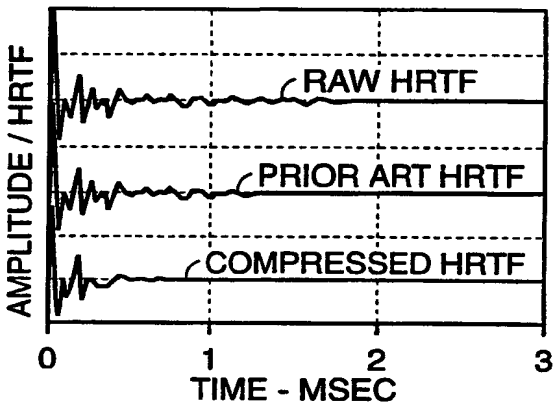


FIG._5a

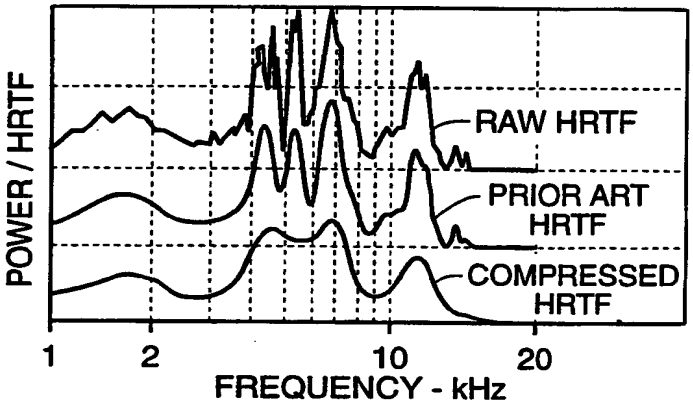


FIG._5b

FIG._6a

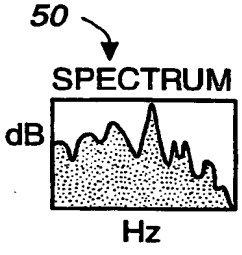


FIG._6b

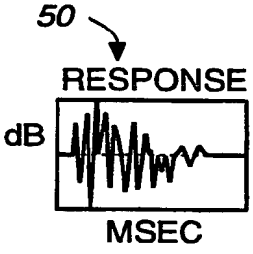


FIG._6c

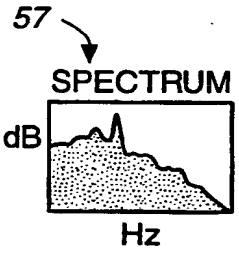


FIG._6d

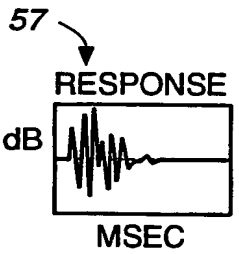


FIG._6e

FIG._7a

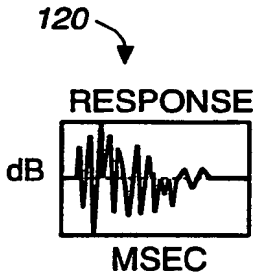


FIG._7b

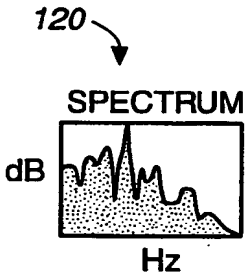


FIG._7c

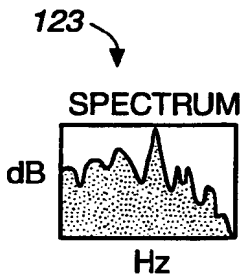


FIG._7d

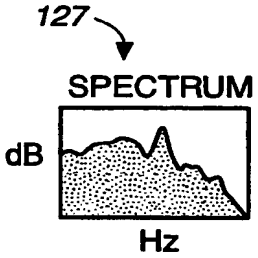


FIG._7e

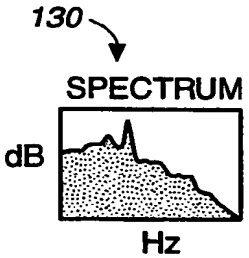


FIG._7f

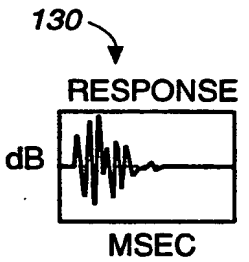


FIG._7g

BNSDOCID: <WO___9531881A1_I_>

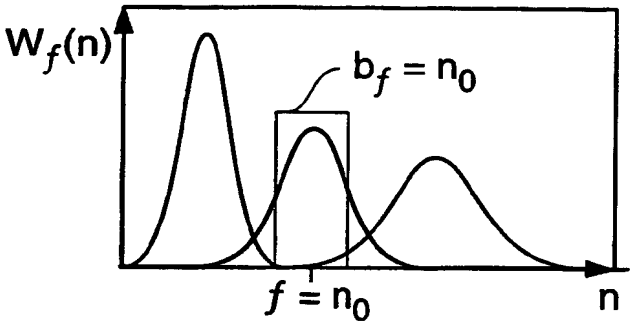


FIG._8

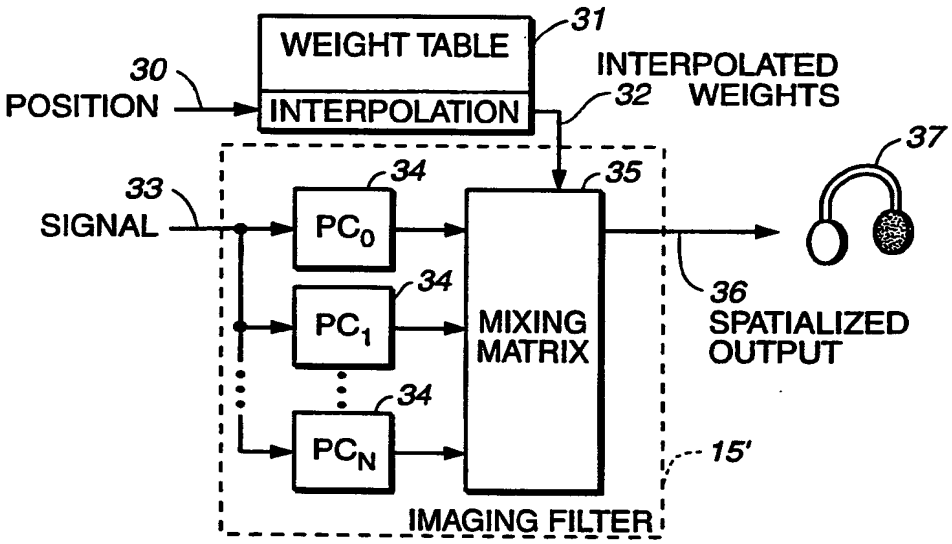


FIG._9

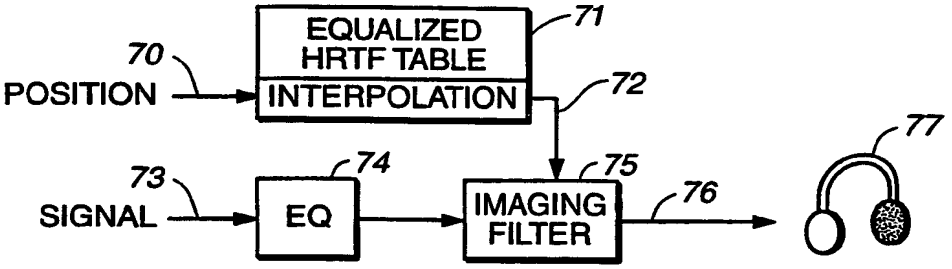


FIG._10

BNSDOCID: <WO___9531881A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/04839

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :H04S 5/00
US CL :381/17,18,25

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 381/17,18,25

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, E	US, A, 5,440,639 (SUZUKI ET AL.) 08 August 1995, see Col. 22, lines 40-41, see Col. 22, lines 64-68.	1-40
Y, E	US, A, 5,438,623 (BEGAULT) 01 August 1995, see figure 1, see Col. 4, lines 29-38, see Col. 12, lines 8-16.	7-8
Y, E	US, A, 5,404,406 (FUCHIGAMI ET AL.) 04 August 1995, see Col. 20, lines 21-29, see Col. 23, lines 17-19 and 30-34.	32-39 and 40

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 SEPTEMBER 1995	Date of mailing of the international search report 04 OCT 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer B. Nguyen for DUC NGUYEN Telephone No. (703) 308-7527

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO___9531881A1_I_>

This Page Blank (uspto)